Kaufmann (2001). However, this requires running tests for each individual sub-region by setting different significance levels. This would lead to higher computation costs.

[0086] For the procedure of selecting the segmentation and combination of regions described above, some validation samples in the opposite class (counterfeits) are needed to evaluate the performance of the selected segmentation and combination. However, compared with the amount of counterfeits that would be required to build a two-class classifier, small amounts of counterfeit validation samples suffice for this validation purpose.

[0087] In order to compare the effectiveness of the one-class classifiers proposed, various tests have been done on both whole notes and segmented notes. Using 3324 genuine notes as a reference set, based on a single feature vector defining the whole note, the D^2 and Mixture of Gaussians (MoG)+Bootstrap tests were employed on 798 counterfeit notes by setting different critical values, that is test significance levels, $\alpha=0.01$, $\alpha=0.05$ and $\alpha=0.10$ to specify the expected FP rate for each test. To obtain the FP rate for genuine notes 10-fold cross-validation was employed. Test results for both tests are listed in **FIG. 3**. It can be seen that the semi-parametric test for all significance levels provides significantly lower levels of false negatives (counterfeits accepted as genuine) than that obtained by the D^2 tests.

[0088] For comparison simple binary classifiers or two-class classifiers have also been tested, specifically estimating the class conditional densities based on both a Gaussian and Mixture of Gaussian models. In this experiment, 3324 genuine and 798 counterfeit were used to train the two-class discriminators. Test results are shown in **FIG. 4**, where 10-fold cross-validation is used for both classes to calculate the values of FN and FP. Employing simple binary classifiers can provide an improvement of around 4% in FN rates over the D^2 test, whilst the difference in performance for the MoG+Bootstrap tests is less pronounced. However, employing the one-class classifiers it is possible to reduce the costly FN rate by increasing the significance level of the test and so increase the number of genuine notes rejected (FP). Clearly in cases where large numbers of examples of counterfeits are available then employing binary classifiers will be advantageous, however one-class classifiers are most useful where the numbers of examples from both classes is highly imbalanced such that the distribution of only one class can be reliably estimated.

[0089] The results set out in **FIGS. 5 and 6** are based on an analysis of the entire note. To test the effect of segmenting the note various segmentations have been tried. In one example R_{max} and C_{max} were set to be **(3,3)**, **(7,7)** and **(15,15)**. Of course, it will be appreciated that R_{max} and C_{max} do not have to be identical they can be any positive integer greater than zero. Using these values optimized note segmentation and classifier combinations were determined, using individual D^2 tests at 1% significance levels. In the experiments the GA used 50 chromosomes in each population. A sample of 2500 genuine notes was used as the reference set for the D^2 test. The GA was trained using 10-fold cross-validation on the 2500 genuine samples as well as 300 counterfeit examples. Finally, tests were performed on an independent test set of 824 genuine notes and 498 counterfeits. The reported test results, shown in **FIG. 7**, are the best selected from 200 repeated runs of the GA.

[0090] The optimized note segmentation and classifier combination strategies obtained by running the GA using the D^2 test are shown in FIGS. **8** to **10**. In comparison to the best results obtained by considering the note as a whole the segmentation and combination strategy has reduced the number or counterfeits wrongly accepted (FN) from 5.76% to 2.41% whilst also significantly reducing the number of genuine notes being erroneously rejected (FP) from 10.1% to 3.03%. Regions of both high contrast with little detail and regions with high levels of detail are combined. Intuitively it would seem likely that the high contrast regions help to reduce the number of genuine notes rejected whilst the highly detailed regions assist in reducing the number of counterfeits wrongly accepted—as these are the regions which are typically more difficult to copy successfully.

[0091] The test described above with reference to FIGS. **8** to **10** was repeated using the mixture of Guassian and bootstrap test. Test results (the best selected from 200 repeated runs) are shown in **FIG. 11**. The optimized note segmentation and classifier combination strategies obtained by running the GA for the mixture of Guassian and bootstrap test are shown in FIGS. **12** to **14**. In this case, when segmenting the note into 4×11 regions, the best performance for the training samples is achieved, whilst the best performance for the independent test set is achieved by segmenting the note into 7×3 regions. This indicates that a more complex segmentation will not necessarily achieve better performance on an independent test set and so a level of 'overfitting' can be observed. In comparison to the best D^2 test results obtained, the best MoG+Bootstrap provides a lower FP value, i.e. number of genuine notes rejected, with a marginal increase in the number of FN's. The decrease in FP value is due to the reduced number of regions being tested (two) and so only two tests at the 1% significance level are being performed. In contrast three regions are combined in the D^2 test and so the expected FP rate is approximately 3× the individual significance levels. Considering the segmentations and combinations again similar regions of low detail and high detail content have been selected as in the previous case.

[0092] Because the GA is a stochastic optimization algorithm, it yields different results in different randomly initialized runs. The GA can be run N_r times and the note segmentation and classifier combination strategy that achieves the best performance can be selected. To investigate the level of solution variability, the GA using mixture of Gaussian and bootstrap was repeatedly run $N_r=200$ times. For the search space **(3,3)**, only two different results (segmentation and combination strategies) were obtained during 200 randomly initialized runs. These are shown in **FIG. 15**. For the search space **(7,7)** thirty-four different segmentation and combination strategies were obtained from the 200 runs. Among them, the best strategy was a 7×3 segmentation of the note and then combining classifiers D_{12} and D_{13}. This achieved a training performance: FN=3.67%, FP=1.88% and a testing performance: FN=2.61%, FP=1.46%. The inter-quartile range (IQR) of values of the training FN, FP and testing FN, FP for those **34** strategies over all runs are respectively: 1.67%, 1.04% and 1.61%, 3.03%. IQR computes the difference between the 75^{th} and the 25^{th} percentiles of the sample. It is a robust estimate of the spread of the data, since changes in the upper and lower 25% of the data do not affect it. On observation the distribution of FN and FP is not normal, IQR is more representative than the standard devia-